Sustainable Collaborator Recommendation Based on Conference Closure

Wei Wang, Jiaying Liu, Zhuo Yang, Xiangjie Kong, *Senior Member, IEEE*, and Feng Xia, *Senior Member, IEEE*

***Abstract*—It is always difficult for researchers to find a new collaborator due to the academic information overload. Although several scientific collaborator recommendation systems have been proposed, most existing methods cannot meet scholars' requirements comprehensively due to the neglect of sustainable collaborations in their schemes. Studies have shown that a sustainable collaboration has a significant positive impact on the productivity and influence for a researcher. To address this problem, we develop a sustainable collaborator recommendation (SCORE) system by utilizing the weak tie relationships brought by academic conferences for SCORE. Through defining and quantifying the conference closure, we incorporate the conference coattending relationships into collaborator recommendation system inspired by the principles of "diversified recommendation." The experimental results on attendees of ten academic conferences show that the SCORE outperforms state-of-the-art collaborator recommendation systems in accuracy and sustainability. Our model can be used to improve the sociability of academic conferences by recommending sustainable collaborators to conference attendees.**

***Index Terms*—Academic information retrieval, collaboration recommendation, scientific collaboration.**

I. Introduction

SCIENTIFIC collaboration [1]–[6] is becoming more and more important in academia. This is because a single scholar may not possess all the expertise to tackle a complex scientific issue. However, it is always difficult for researchers to find a new collaborator due to the academic information overload. Many scientific collaborator recommendation systems have been proposed to solve such a problem. These systems mainly take the advantages of the theory of link prediction which aims to predict whether two nodes will connect with each other or not [7], [8].

Manuscript received August 3, 2018; revised November 20, 2018; accepted January 31, 2019. Date of publication March 1, 2019; date of current version April 1, 2019. This work was supported in part by the National Natural Science Foundation of China under Grant 61872054 and in part by the Fundamental Research Funds for the Central Universities under Grant DUT18JC09. *(Corresponding author: Xiangjie Kong.)*

W. Wang is with the Key Laboratory for Ubiquitous Network and Service Software of Liaoning Province, School of Software, Dalian University of Technology, Dalian 116620, China, and also with the Department of Computer and Information Science, Faculty of Science and Technology, University of Macau, Macau 999078, China.

J. Liu, Z. Yang, X. Kong, and F. Xia are with the Key Laboratory for Ubiquitous Network and Service Software of Liaoning Province, School of Software, Dalian University of Technology, Dalian 116620, China (e-mail: xjkong@ieee.org).

Digital Object Identifier 10.1109/TCSS.2019.2898198

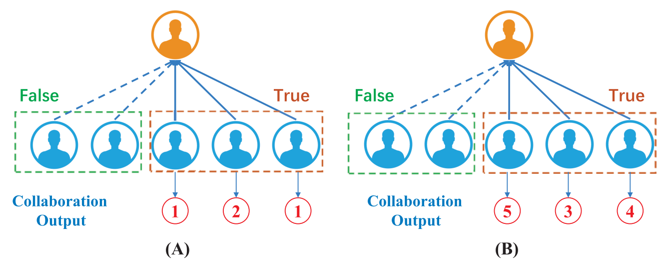


Fig. 1. Comparison between regular and SCOREs. Systems A and B have the same precision (60%), whereas the average CO of true scholars recommended by B is higher that A (4 > 1.33).

However, most existing collaborator recommendation algorithms only concentrate on the predicted link, by which the collaborators are selected and recommended [9]. As we know, in reality, the scientific collaboration is not a one-shot deal. The sustainability of different collaborations may have great differences. Here, the sustainability denotes the persistence of a collaboration. Specifically, we explore the collaboration sustainability from three perspectives, including collaboration output (CO), collaboration duration (CD), and collaboration index (C-index) (see Section III-A).

Previous research [10] demonstrates that scholars will benefit from a sustainable collaborator in productivity and innovation. But, it is more difficult to identify a sustainable collaborator. Thus, when designing a collaborator recommendation system, we should not only consider the precision but also the sustainability. For example, in Fig. 1, two collaborator recommendation systems have the same precision (60%). However, system B obviously outperforms system A because its recommended scholars have more CO with target scholar.

Despite the importance of sustainable collaborators, no previous work has studied how to find and recommend sustainable collaborators. Therefore, in this paper, we propose and study the problem of sustainable collaborator recommendation (SCORE). However, there are two challenges to address the problem. First, in order to enhance the sustainability, we have to identify potential sustainable collaborators based on mechanisms of sustainable collaborations. In previous studies, the collaborator recommendation systems are mainly designed based on the social rule of triadic closure [11], [12] which assumes that two people are possibly to be connected if they share a common friend. However, these recommendation systems overlook the strength of weak ties [13] where the majority of the novel information dissemination is generated by weak ties. Second, by assuming a reliable weak tie relationship for sustainable collaborations, how can we effectively incorporate such knowledge into existing recommendation systems to improve the recommendation quality.

To tackle both challenges mentioned above, we propose an SCORE system. The SCORE model is designed based on the theory of conference closure which denotes that two scholars will connect with each other if they have attended the same conference [14], [15]. Collaborators meeting by accident in an academic conference are important and serendipitous, who have higher novelty [16]. Moreover, they have similar research interests. Thus, such collaborations will last longer.

Since the random walk with restart model has been proven to be effective in many similarity-based recommendation systems [8], [17], we adopt it as our basic model. We utilize the conference closure to connect scholars of weak tie relationships to bias the random walk model in collaboration networks for SCORE. Through extensive experiments on attendees of ten academic conferences in DBLP digital library, we demonstrate that our proposed SCORE model can recommend more sustainable collaborators than the baseline model with losing accuracy. Our main contributions can be summarized as follows.

1) *Problem Formulation:* We investigate SCORE problem, targeting significant improvement in collaborator recommendation quality, meanwhile maintaining accuracy. Then, we define three factors including CO, CD, and C-index to measure the collaboration sustainability.
2) *Recommendation Algorithms:* We propose SCORE, a novel collaborator recommendation system, where the weak tie relationship, i.e., conference closure is utilized to bias random walk for SCORE.
3) *Experimental Design:* We conduct extensive experiments on attendees of ten academic conferences to verify the effectiveness of SCORE for SCORE by comparison with the state-of-the-art collaborator recommendation methods.

Organization: The rest of this paper is organized as follows. Section II provides a brief survey of related work. Section III introduces some preliminary knowledge related with our work including the sustainable collaboration and conference closure. Section IV proposes the SCORE algorithm for SCORE. The experimental results are illustrated in Section V. Section VI concludes this paper.

II. Preliminaries

A. Sustainable Collaboration

Scientific collaboration is usually not a one-shot deal [18]. Scholars may collaborate with each other more than once. Previous study has proven sustainable collaborators will benefit scholars in publication quantity and academic impact [10]. In Fig. 2, we show the distribution of all the collaboration of DBLP data set in terms of CO and CD. We can observe from Fig. 2 that the most collaborations last for a short-time period and few times. Considering the importance of

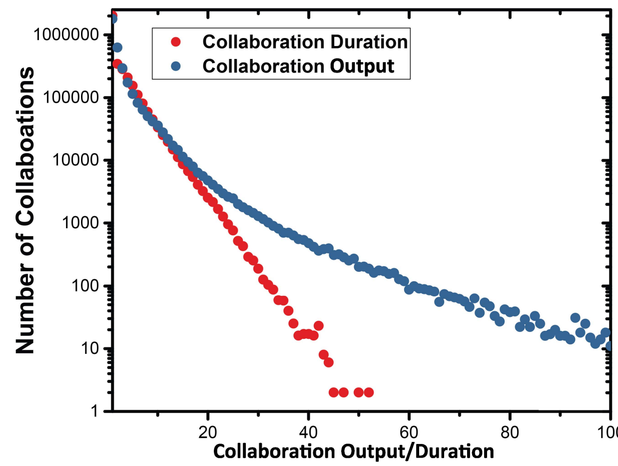


Fig. 2. Distributions of the CO and duration.

sustainable collaborators, we need to recommend collaborators to scholars with longer CD and more COs. Specifically, in this paper, we explore the sustainability of scientific collaboration from three perspectives, including CO, CD, and C-index:

1) Collaboration Output: CO is the direct reflection of a collaboration sustainability, which shows how many papers two scholars have coauthored with each other. In most cases, a sustainable collaborator will bring a higher CO. However, in some extreme cases where two scholars coauthor many papers (e.g., 15 papers) in a short-time period (e.g., two years), those collaborations are not sustainable collaborations.

2) Collaboration Duration: CD reflects how long two scholars have collaborated with each other. Usually, a sustainable collaboration will last for a long time. But, in some extreme cases, where two scholars collaborate for a long time (e.g., 15 years) with few publications (e.g., two papers), such collaborations cannot be regarded as sustainable collaboration neither.

3) Collaboration Index: C-index is proposed to better quantify the collaboration sustainability. The C-index is inspired by the idea of h-index which is proposed to quantify scientific output [19]. Specifically, a collaboration has C-index h if two scholars collaborate with each other at least h times within at least h years. For example, the collaboration between scholars A and B has ten C-index if they collaborate with each other at least ten times during ten years.

B. Conference Closure

The theory of triadic closure [11], [12], [20] has been proposed to explore the mechanism of link formation. The triadic closure indicates that if two people share a common friend, these two people are possibly to be connected in the future. The famous friend recommendation system "People You May Kno" in online social media is designed based on this theory [21]. Based on the triadic closure, people may choose new acquaintances who are friends of friends [12]. Meanwhile, the focal closure (or membership closure) extracted from affiliation networks assumes that two scholars may collaborate with each other if they belong to the same institution [22].

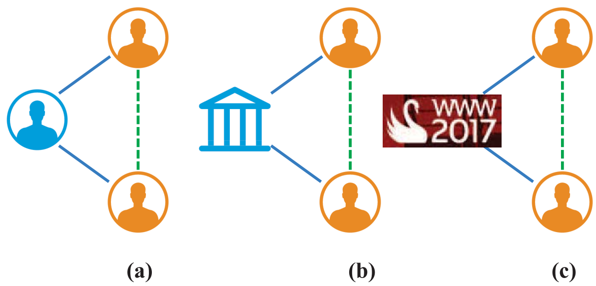


Fig. 3. Examples of (a) triadic closure, (b) focal closure, and (c) conference closure.

The examples of the triadic closure and focal closure are shown in Fig. 3(a) and (b), respectively.

Compared with the academic journals, the academic conference is not only a venue for publications but also a nursery room for new scholar encounters. Academic conferences play an important role in bringing scholars together for idea sharing, work discussion, and experience communication. Based on the triadic closure, focal closure, and the social functions of academic conferences, the conference closure has been proposed which assumes that two scholars are possibly to be connected if they attend the same conference [15]. The example of the conference closure is shown in Fig. 3(c). On the one hand, the basic assumption of these three theories is that individuals tend to interact with others who are similar to them, which is known as homophily [23]. On the other hand, the differences between the conference closure and the other two closures are obvious because of the characteristics of academic conferences. Academic conferences usually are held once a year, which makes the conference closure more like an event-based relationship. Most importantly, scholars attending the same conference may share the same research topic and have read each others' publications before, whereas they may be totally strangers to each other in reality. Thus, the connections brought by the conference closure can be regarded as a kind of weak tie [13]. Such weak ties may benefit scholars in access to novel information and resources by bridges (connections outside their circle of acquaintances) [13] or by spanning structural holes [24].

In order to incorporate the conference closure into SCORE, we first need to quantify the conference closure of each conference. In fact, although there are many qualitative analysis of the triadic closure and focal closure, no quantization methods has been proposed. In this paper, we follow our previous study on conference closure [14], [15] to quantify the conference closure. Specifically, based on the study in [14] and [15], the conference closure of a given scholar i at the conference C is calculated as

$$\xi_i = \frac{C_i^c}{C_i^{\text{all}}} \tag{1}$$

where C_i^c is the number of new collaborators resulted from the conference closure of attendee i and C_i^{all} is the total number of new collaborators of attendee i. The conference closure ξ_i can reflect the impact of the conference on scholar i's choice of new collaborators. Meanwhile, the conference closure of

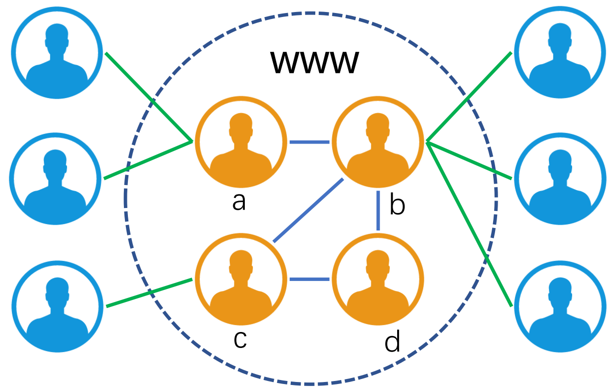


Fig. 4. Example of conference closure quantization.

the conference C is the average of all attendees' ξ, which can be calculated as

$$\xi_C = \sum_{i}^{m} \frac{\xi_i}{m} = \sum_{i=1}^{m} \frac{C_i^c}{mC_i^{\text{all}}} \tag{2}$$

where m is the number of attendees of the conference. Note that the conference closure is calculated in the next following one–five years after the conference was held.

Fig. 4 shows an example of the conference closure quantization. In Fig. 4, we assume that the total attendee number m of the conference WWW is 4 including scholars a, b, c, and d. The conference closure ξ_i of each scholar is 0.33, 0.5, 0.66, and 1, respectively. Thus, the conference closure of WWW ξ_{WWW} is 0.625 in this example. Actually, the conference closure of practical conference is much lower than the value in this example. We have given the exact conference closure of investigated conferences in Table II. Finally, the quantified conference closure is used to build a new connection between scholars who have not collaborated with each other but attended the same conference to enhance sustainability.

III. Design of SCORE

A. Problem Statement

The problem we investigate in this paper is to propose an SCORE in scientific collaboration network for a given scholar. The sustainable collaboration means that two scholars will collaborate more than once with each other. While previous work in scientific collaborator recommendation mainly predicts whether two scholars will collaborate with each other or not, we focus on the SCORE. In reality, scientific collaboration is not a one-shot deal. Thus, when we recommend collaborators, we should not only consider precision but also the recommendation diversity, which is, in this paper, sustainability. We take an advantage of conference closure for SCORE based on the theory of serendipity suggestion [25]: items should be not only relevant and novel to the target user but also obviously different from the items that the user has rated. An academic conference is not only a venue for publication but also a nursery room for new scientific encounters. We assume that collaborations born from conference closure may be more sustainable.

TABLE I
LIST OF VARIABLES

Symbol	Definition
$scholar\ i$	Scholar attending academic conferences
$scholar\ x$	Target scholar for recommendation
$scholar\ y$	Candidate for recommendation
n_i	Node n_i in the collaboration network
S_{xy}^{RWR}	RWR similarity between scholar x and y
MR	$N * 1$ ranking score vector
$L(n_i)$	Number of all the neighbors of node n_i
$M(n_i)$	Set of nodes incident to node n_i
α	Damping coefficient
N	Total number of nodes in a graph
$\mathbf{S}$	Transfer matrix
q	$N * 1$ initial status of π
t	Iteration times
$L_{weight}(n_i, n_j)$	Link weight of n_i and n_j
$c(n_i, n_j)$	Conference closure between n_i and n_j
$k(n_i, n_j)$	Collaboration output between n_i and n_j

Generally, the SCORE can be formulated as follows.

Given a snapshot of a scientific collaboration network at time T and the publication records in digital libraries of a target scholar x.

Recommend a list of sustainable potential collaborators $\{y_1, y_2, y_3 \dots y_n\}$ based on the similarity $score\langle x, y\rangle$ considering conference closure. The target scholar x will get connected with potential collaborators in the near future time $T' = T + \Delta T$.

To solve this problem, we first extract the publication records from the digital libraries during a past time interval $T_{\text{whole}} = [t_{\text{start}}, t_{\text{end}}]$ and split it into two parts including $T_{\text{train}} = [t_{\text{start},t_{\text{recommend}}}]$ and $T_{\text{test}} = (t_{\text{recommend}}, t_{\text{end}}]$. The data in the first time interval T_{train} is used to construct the collaboration network, and recommend the potential collaborators in the second time interval T_{test}. The publication records in the second time interval T_{test} are used as the ground truth for evaluating the performance of the recommendation systems. It is worth mentioning that the goal of our recommendation system is to recommend new collaborators considering sustainability, rather than the coauthored ones, to the target scholars using the weak ties brought by conference closure.

B. Model Description

Similar to previous work in link prediction, the SCORE model recommends collaborators based on the idea that two similar scholars are possibly to be connected in the future [26]. In order to calculate the similarity between two scholars in scientific collaboration networks, we employ the random walk model with restart (RWR) model [17] as our foundation. It has been proven to be effective in many scientific collaboration systems. Meanwhile, we further introduce link weight information including conference closure and CO to bias the random walk so that both the similarity and serendipity can be considered. Table I summarizes notations used in the SCORE model. The framework of SCORE is shown in Fig. 5.

In SCORE model, the sustainable collaborator is recommended based on the similarity of other scholars to the target scholars. Based on these two metrics, each recommended scholar y has a rank score S_{xy}^{RWR} to the target scholar x, which can be described as

$$S_{xy}^{\text{RWR}} = \pi_{xy} + \pi_{yx} \tag{3}$$

where π_{xy} (or π_{yx}) denotes the importance of scholar y (or x) to scholar x (or y). π_{xy} is determined by two factors, the number of scholars connected to scholar x and the weight of these scholars. It can be calculated by

$$\pi_{xy} = \frac{1-\alpha}{N} + \alpha \sum_{n_j \in M(n_y)} \frac{\pi_j}{L(n_j)} \tag{4}$$

where α is the probability that the walker will continue walking to next neighbor, $L(n_j)$ is the number of all the neighbors of scholar n_i, and π_j is the rank score of scholar n_j to target scholar n_i. The π_{yx} is calculated similar with π_{xy}. In the SCORE model, the walker randomly walks to next node. Equation (4) merely shows the final step to get the rank score of a node. The whole model is an iterative process, which can be described as follows:

$$\pi^{t+1} = \alpha \mathbf{S} \pi^t + (1-\alpha) q \tag{5}$$

where $\mathbf{S}$ is the transfer metric of the network and q is the initial status. Based on this equation, the initial state π^0 is q, which is a row vector.

1) Link Weight: Previous random walk models usually assume that the weight of every link is the same and the entry of matrix $\mathbf{S}$ is calculated as $\mathbf{S}_{ij} = 1/L(n_j)$. In order bias the walker to sustainable collaborators resulted from conference closure, we define the matrix $\mathbf{S}$ with link weight based on CO and conference closure. The link weight $L_{\text{weight}}(n_i, n_j)$ mainly contains two features including conference closure $c(n_i, n_j)$ and CO $k(n_i, n_j)$.

a) Conference Closure: In order to bias SCORE model to more sustainable potential collaborators, we utilize the weak ties derived from conference closure. The link weight brought by the conference closure is the quantified conference closure ξ_C of the conference, which is calculated in (2). Meanwhile, two stranger scholars may attend c common conferences. Thus, the conference closure relationship $c(n_i, n_j)$ between n_i and n_j is calculated by

$$c(n_i, n_j) = \sum_{\lambda=1}^{c} \frac{\xi_{C_\lambda}}{c} \tag{6}$$

where C_λ denotes the conference closure of each conference.

b) Collaboration Output: Two scholars may collaborate with each other more than once. It is necessary to consider the number of CO when calculating the link weight between two scholars. The CO $k(n_i, n_j)$ between two scholars is calculated as

$$k(n_i, n_j) = k \tag{7}$$

where k denotes the number of times two scholars have coauthored with each other before.

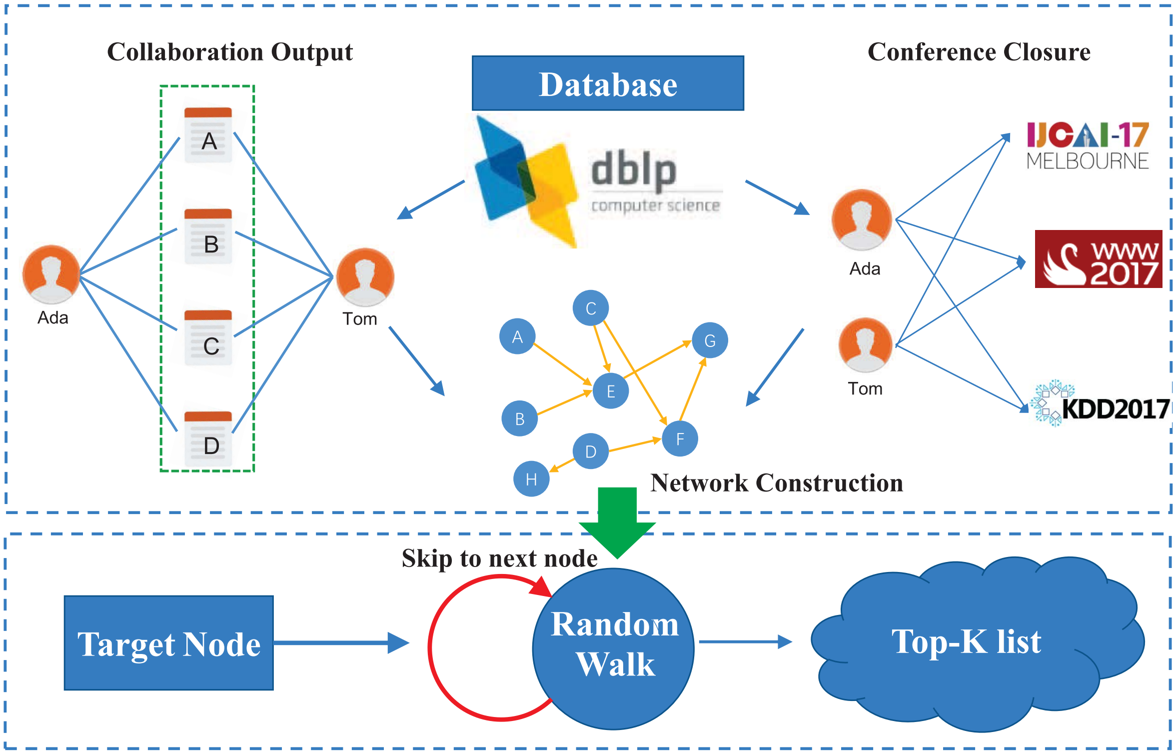


Fig. 5. Framework of SCORE.

Algorithm 1 SCORE (R, a, MaxIteration, MinDelta)

Input: Conference clsoure relationships between node pairs, $c(n_i, n_j)$; Collaboration times between node pairs, $k(n_i, n_j)$; Collaboration networks, G=(N, L)

Output: A list of potential collaborator;

$\mathbf{S} \leftarrow$ ComputeTransferMatrix()
$MR_0 \leftarrow R$
$Q \leftarrow R$
for $k \leftarrow 0$ to $MaxIteration - 1$ **do**
  $diff \leftarrow 0$
  for $i \leftarrow 0$ to $len(Q) - 1$ **do**
    $MR_{k_i} = \alpha \sum_{j=0}^{len(Q)} S_{i,j} MR_j + (1-\alpha) Q_i$
    $diff \leftarrow diff + (MR_k - MR_{k-1})$
  end for
  if $diff < MinDelta$ **then**
    break
  end if
end for
$Predictions \leftarrow predictions(MR)$
return $Predictions$

2) Workflow: While the transition matrix is different, we use the same random walk algorithm with [8] and [27]. Specifically, the procedure of the SCORE shown in Algorithm 1 can be described as follows.

1) *Network Construction:* We first construct the scientific collaboration network based on the coauthor records extracted from the DBLP digital library. In scientific collaboration networks, the nodes are the scholars and two scholars are connected if they have coauthored at least one paper.
2) *Link Weight Calculation:* The link weight between two scholars is calculated based on the CO and conference closure. Specifically, the $L_{\text{weight}}(n_i, n_j)$ is calculated by
$$L_{\text{weight}}(n_i, n_j) = c(n_i, n_j) + k(n_i, n_j) \tag{8}$$
$$= \sum_{i=1}^{c} \sum_{i=1}^{m} \frac{\xi_{C_i}}{c} + k \tag{9}$$
where m denotes the number of conference attendees, c is the number of conferences both of the two scholars attend, and k denotes the number of times two scholars have coauthored with each other before.
3) *Random Walk:* Before the random walk, we first acquire the transfer matrix $\mathbf{S}$ with weighted links. Then, SCORE model starts with initializing the rank score π^0 and restart status q as (0,…,1,…,0). The target node n_i is set as 1 and the other nodes are set as 0. The SCORE model walks based on (5). Thus, we can get the rank score vector π_{xy}.
4) *Recommendation:* Finally, we sort scholars based on their corresponding rank scores to target scholar. The Top-N scholars are recommended to the target scholar. Meanwhile, we take out those scholars who have collaborated with target scholars before.

C. Discussion on Conference Closure

As shown in Fig. 6, the biggest advantages of SCORE recommendation model is that it could improve the recommendation diversity by incorporating potential collaborators

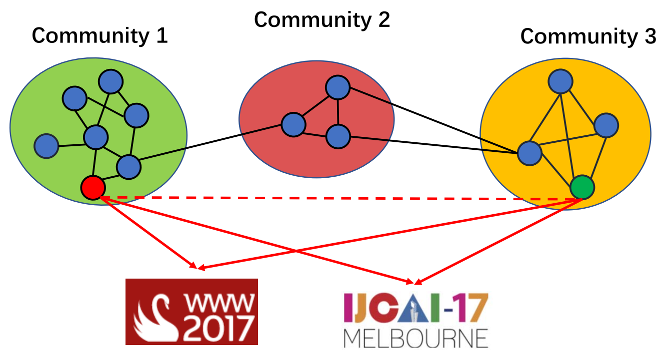


Fig. 6. Advantages of SCORE recommendation. The greed node can be recommended to the red node with higher possibility because of the direct link resulted from conference closure.

TABLE II
STATISTICS OF INVESTIGATED CONFERENCES

Conference	Field Rating[1]	ξ_C[2]
SIGIR	89	0.078
SIGMOD	97	0.081
WWW	83	0.068
VLDB	77	0.063
KDD	122	0.088
ICDE	70	0.064
CIKM	65	0.052
ICML	72	0.054
ISWC	62	0.041
NIPS	78	0.065

[1] Field rating is gained from Microsoft Academic Search.
[2] ξ_C is calculated three years after the conference.

resulted from conference closure. In Fig. 6, if we recommend collaborators for the target node (red node) with basic RWR model, the recommended scholars are mainly from community 1. Such collaboration may last for a short-time period with few collaboration times. The system can barely recommend the green node to the target scholar. However, if the red node begins collaborating with the green node, their collaborations will last longer. The reason behind this phenomenon is that these two scholars may benefit from the weak tie based on conference closure [15]. In the SCORE model, such weak relationships can be effectively utilized to enhance diverse and SCORE.

IV. Experimental Results

In this section, we evaluate our proposed model based on ten famous conferences in the field of data mining extracted from DBLP digital library. All the experiments are performed on a 64-bit Windows-based operation system, with a 4-duo and 2.6-GHz Intel Xeon CPU, 128-GB memory. The experiments are operated with python.

A. Experimental Setup

1) Data Set: It is difficult to get the participant list of various conferences, in order to calculate the conference closure. Hence, we assume that the first author of a conference paper as the participant of the conference. The reason is that, in most cases, at least one author should register and attend the conference and the participant usually is the first author of the paper. Meanwhile, the first author may have the highest contribution of the paper. We extract all the publication records from DBLP [28] in 2000 to construct the scientific collaboration network. Here, the "new" collaborations mean two authors who had never collaborated with each other before attend a same conference with individual papers, and then these two authors have a joint paper afterwards. We traverse all previous digital records to detect whether two scholars had cooperated before, which is not limited to the single year of 2000 in DBLP. The ground truth of collaboration relationships are collected in the next ten years. The target scholars are extracted from ten academic conferences. These conferences are in the field of data mining, including ACM Special Interest Group on Information Retrieval, ACM Special Interest Group on Management Of Data, International World Wide Web Conferences, International Conference on Very Large Data Bases, ACM Special Interest Group on Knowledge Discovery and Data Mining, International Conference on Digital Entrepreneurship, ACM International Conference on Information and Knowledge Management, International Conference on Machine Learning, International Symposium on Wearable Computers, and Annual Conference on Neural Information Processing Systems. These conferences are the best conferences in the field of data mining and all of them are held once a year. The publications accepted by these conferences are highly selected via a peer review process. Meanwhile, at least one author of each accepted paper should register and attend the conference. Usually, the attendees should give a talk about their papers to share their ideas and communicate with peers. We then extract their publication records in the next ten years as the test set for evaluation. The data set contains 65 587 scholars and 111 826 links. There are 1156 conference participants (target scholars).

Note that all these conferences are from the field of computer science. It is known that in many other scientific fields, academic conferences are also important. Although they play a somewhat different role than in computer science, conferences in other fields also are nursery rooms for new scholar encounter [15]. The benefit of such offline communication from attending conference in all fields will help find related and unexpected collaborators. Due to the data limitation, we can not evaluate the performance of SCORE on conferences from other fields. But, we have confidence in SCORE over other fields because once the collaboration resulted from conference closure is established, it will be more sustainable. The reason is that the motivation of beginning such collaboration is stronger and such chance is relative rare.

The statistics of the investigated conferences are shown in Table II. Specifically, the conference closure of each conference is calculated three years after the conference. The reason is that, it takes time for coauthored paper to get published. We can see from this table that, conferences with high academic field rating will have higher quantified conference closure. Meanwhile, the conference closure ξ_C of any conference is relatively small (less that 10%) [15].

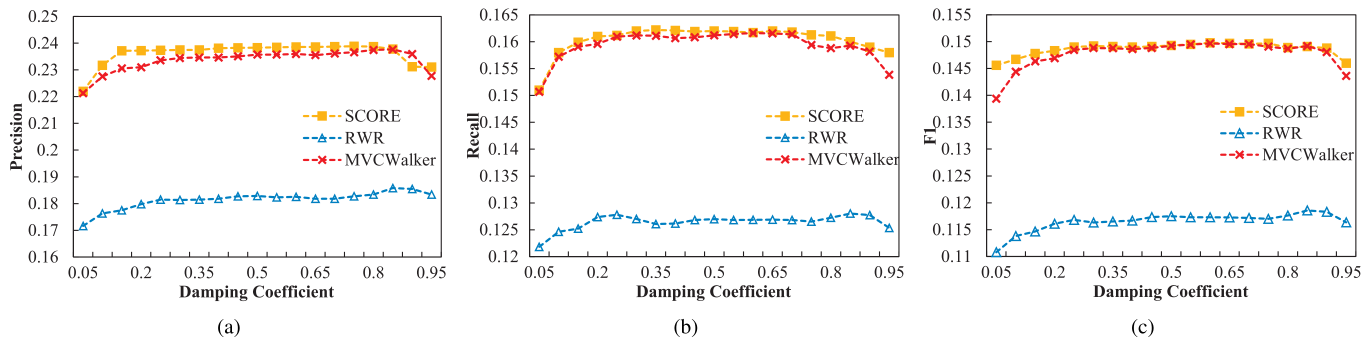


Fig. 7. Comparison between SCORE and baseline methods in terms of accuracy over damping coefficient. (a) Precision. (b) Recall. (c) F1.

Thus, the weak tie relationship brought by conference closure is smaller than that of direct collaboration relationships.

2) Evaluation Metrics: In order to evaluate the performance of our proposed model, we take advantages of historic collaboration relationships (2000) for recommending and the next ten years (from 2000 to 2010) for evaluation. Since our recommendation system SCORE aims to recommend accurate and sustainable collaborators to target scholars, we evaluate our proposed model and baseline models from the perspectives of accuracy and sustainability.

a) Accuracy: In this paper, we utilize three basic accuracy metrics in terms of precision, recall, and F1 [29], [30].

b) Sustainability: In Section II, we have introduced the definition of sustainable collaborators from three perspectives, including CO, CD, and C-index. During the experiments, we adopt these three metrics to evaluate the sustainability of recommended collaborators of SCORE and baseline methods.

3) Baseline Method: In the experiments, we compare our proposed SCORE model with two random-walk-based methods including basic RWR and MVCWalker.

a) RWR: RWR is the basic random walk model without considering the link weight as well as the conference closure relationship.

b) MVCWalker: MVCWalker [8] is proposed based on RWR model. This method considers three academic factors to weight links between scholars, i.e., coauthor order, latest collaboration time, and CO. The goal of this model is recommend most valuable collaborator (MVCs) to target scholars.

c) PCJS: PCJS is a learning-based method for prediction collaboration for junior scholars. It takes advantages both network features and proposed features including affiliation, sting distance, geographic distance and content similarity. We gain the institution information from Aminer [31]. The content similarity are calculated based on the title and keywords.

B. Accuracy Comparison

1) Effect of Damping Coefficient: The damping coefficient plays an important role in random-walk-based recommendation system. Based on (5), the damping coefficient determines the probability for the walker to jump back to the starting node (target scholar). This parameter has a significant impact on the performance of SCORE since it controls how fast the random walker will converge. Thus, we investigate the performance of SCORE and baseline methods with different damping coefficient ranging from 0.1 to 0.95.

We evaluate the effect of damping coefficient on the performance of SCORE and baseline methods in terms of precision, recall, and F1, respectively. The PCJS is not used because it is learning-based without damping coefficient. The results are shown in Fig. 7. We can see that with the increase of damping coefficient, SCORE, RWR, and MVCWalker have similar trends in terms of precision, recall, and F1, respectively. Specifically, with the increase of damping coefficient, the precision, recall, and F1 all go up gradually and reach the peak at 0.85. In Fig. 7, we can see that both SCORE and MVCWalker have higher precision, recall, and F1 than that of basic RWR, which implies that considering academic factors such as the conference attending could improve the accuracy of academic collaborator recommendation. Meanwhile, SCORE outperforms both RWR and MCVWalker in terms of precision. However, such improvement is small as the precision, recall and F1 of SCORE and MVCWalker are almost the same. All the accuracy of these three recommendation system have relative low accuracy. The reason is that the ground truth data is directly extracted from the digital libraries, which, to some extent, can not comprehensively reflect the impact of the recommendation system on scholars' choices of candidate collaborators. However, this method could reflect the performance of SCORE model by comparison with baseline methods. We can see that our proposed SCORE outperforms state-of-the-art accuracy with different damping coefficients.

2) Effect of Recommendation List: Fig. 8 depicts how the length of the recommendation list influences the performance of SCORE, RWR, MVCWalker, and PCJS in terms of precision, recall, and F1, respectively. It can be seen from Fig. 8(a) and (c) that with the increase of recommendation list, there has been a marked drop in precision and F1 of all recommendation systems. The reason can be explained based on the definition of the metric precision because the total number of the recommended collaborators increases faster than the number of correct recommendations. The overall precision, recall, and F1 of SCORE are higher than that of RWR, MVCWalker, and PCJS. The increase of recall of all three recommendation system, as shown in Fig. 8(b), can also be explained based on the definition of the metric recall because more correct scholars are recommended and the number of true collaborators remains unchanged.

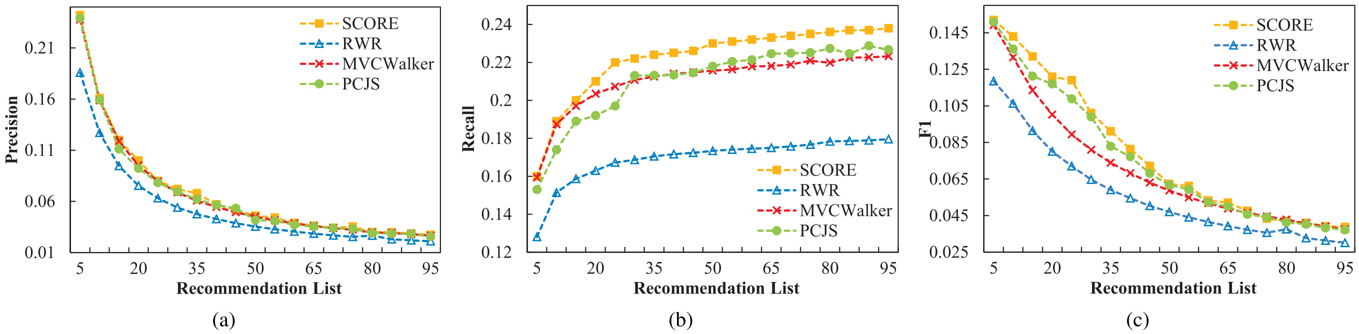


Fig. 8. Comparison between SCORE and baseline methods in terms of accuracy over recommendation list. (a) Precision. (b) Recall. (c) F1.

C. Sustainability Comparison

We set the damping coefficient as 0.85 and the number of recommended collaborators as 20 for the randomwalk-based method and set the parameters of PCJS. We perform 100 times for each model for randomly selected 100 target nodes. The results are shown in Fig. 9. The COs of SCORE, RWR, and MVCWalker are 4.3, 2.4, and 3.1, respectively. In Fig. 9(b), the CD of SCORE is 2.9, in comparison to 2.1 of RWR and 2.3 of MVCWalker. Meanwhile, SCORE has the highest C-index, 2.2 which is higher than 1.7 of RWR and 1.8 of MVCWalker as can be seen in Fig. 9(c). Meanwhile, the C-index of every recommendation system is lower than CO and CD. The reason is that the metric C-index considers both CO and CD which will lead to a low value. We can see from Fig. 9 that the overall performance of SCORE is much better than the state-of-the-art methods in terms of collaboration sustainability, which demonstrates that our proposed SCORE model could recommend more sustainable collaborators.

As shown in Figs. 7–9, and Table III, we can see that the proposed SCORE model recommends more sustainable collaborators and increase accuracy. Usually, there will be a tradeoff between accuracy and diversity (sustainability). However, SCORE utilizes the power of weak tie relationships brought by conference closure can aviod such tradeoff.

D. Conference Comparison

We further show the performance of SCORE and other methods for each investigated conference in Table III. We can observe from this table that are as follows.

1) SCORE has best performance for most conference and the learning-based mehtod PCSJ outperforms RWR and MVCWalker.
2) All methods perform better for high field rating (see Table II) conferences. For example, the precision of SCORE on KDD (31.1%) is larger than that of ISWC (19.3%).
3) The collaboration sustainability including CO and CD resulted from high field rating conference is higher. For example, the CO (4.32) of SCORE on KDD is higher than that of ISWC (3.35). These observation indicate that it is correct to incorporating conference closure for sustainable collaborator recommendationSCORE.

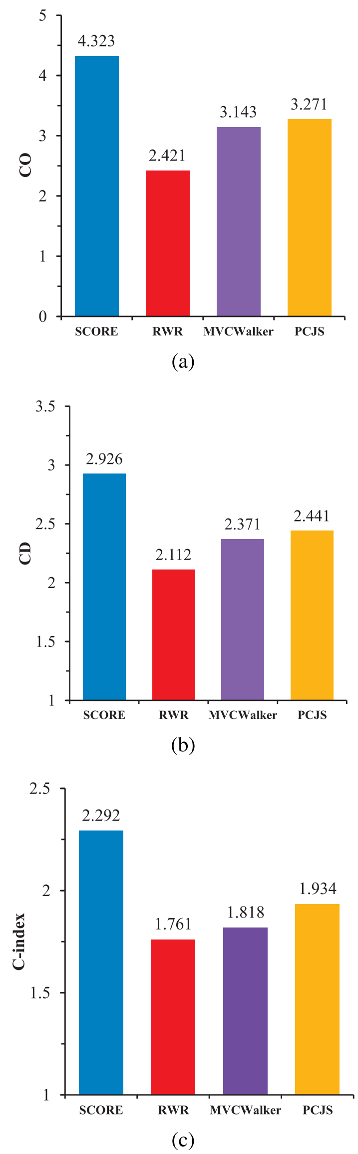


Fig. 9. Comparison between SCORE and baseline methods in terms of collaboration sustainability. (a) CO. (b) CD. (c) C-index.

V. Related Work

In this section, we review some well-known related work in the context of collaborator recommendation, random walk model, and diversified recommendation.

TABLE III

RESULTS COMPARISON OF DIFFERENT CONFERENCES. TO SAVE SPACE, WE USE **S**, **R**, **M**, AND **P** TO STAND FOR SCORE, RWR, MVCWALKER, AND PCJS, RESPECTIVELY. THE BOLD NUMBER INDICATES THE BEST PERFORMANCE OF FOUR SYSTEMS FOR A SPECIFIC CONFERENCE IN TERMS EACH METRIC

Conference	Precision (%)				Recall(%)				CO				CD			
	S	R	M	P	S	R	M	P	S	R	M	P	S	R	M	P
KDD	**31.1**	19.2	24.1	30.8	**22.7**	15.7	21.1	19.8	**4.32**	4.13	4.21	4.07	**3.27**	3.13	3.22	3.08
SGMOD	**28.4**	22.6	23.4	28.1	**20.4**	14.7	18.8	17.8	4.09	3.78	**4.11**	3.99	3.09	3.08	**3.17**	2.89
SIGIR	25.2	21.2	**26.8**	24.8	**21.7**	13.8	20.3	18.2	**4.14**	4.07	3.77	3.73	**3.17**	2.89	2.48	3.01
WWW	28.9	19.4	18.4	**30.2**	**28.6**	11.9	15.8	14.7	**4.27**	3.73	3.58	3.81	**2.93**	2.56	2.70	2.80
NIPS	27.1	26.8	27.7	**28.2**	16.4	12.5	15.5	**16.7**	**4.17**	2.43	3.15	3.54	**2.83**	2.71	2.39	2.52
VLDB	16.2	14.1	17.2	**16.9**	14.3	12.9	**19.4**	17.2	**4.28**	2.89	3.28	3.62	**2.78**	2.18	2.49	2.47
ICML	**23.7**	13.6	25.8	22.7	12.7	13.6	12.6	**16.8**	**3.73**	3.28	2.93	3.27	**2.89**	2.20	2.31	2.81
ICDE	17.2	12.2	16.7	**17.9**	**16.5**	12.1	15.8	14.4	**3.54**	2.91	2.87	3.40	2.58	2.32	2.28	**2.66**
CIKM	**22.5**	15.9	18.8	18.3	12.3	10.7	**13.7**	12.7	**3.61**	2.35	3.08	3.19	2.62	1.89	**2.69**	2.45
ISWC	**19.3**	11.6	14.6	17.3	10.4	9.21	11.7	**12.2**	**3.35**	2.24	2.84	3.01	**2.54**	1.93	2.17	2.35

A. Collaborator Recommendation

Scientific collaborator recommendation has been mainly explored from the perspective of link prediction based on social network analysis [26], [32]–[39]. The link prediction aims at inferring the evolution of a social network using features intrinsic to the network itself [7], [40]. In previous research, the goal is to identify the probability of the future collaborations [8]. In collaborator recommendation, the stranger scholar pair $\langle x, y\rangle$ is assigned a connection weight $score(x, y)$ based on the social network and the recommendation list is given by the decreasing order of score. Thus, it can be regarded as computing the similarity between scholars x and y. However, the collaborator recommendation for scholars is different from link prediction because it is context-based. We need to consider the academic factor when recommending collaborators.

Due to growing tendency of researching towards collaborative research, many recommendation systems have been proposed, including the supervised approach by network-based features [41]–[43], unsupervised approach by citation network [44], and ranking approach by node similarity [8], [40]. Chen *et al.* [41] have developed the CollabSeer system to recommend potential research collaborators for scholars based on vertex similarity including Jaccard similarity, cosine similarity, and relation strength similarity. Supervised methods try to extract multiple features based on network topology and node attributes, and learn to recommend with machine learning methods [45]. Xia *et al.* [8] have proposed MVCWalker model to recommend suitable collaborators based on node similarity by considering several academic factors including collaboration times, CD, and coauthor order. Tang *et al.* [46] propose to recommend cross-domain collaboration by considering the scholars' research topics.

So far, no recommendation approach has been proposed to recommend sustainable collaborators. Previous studies have shown that the sustainable collaborations have a significant positive impact on productivity and citations [10]. Meanwhile, it has been explored that academic conferences can promote new scientific collaboration [15]. Thus, we propose to utilize the conference attending information for SCORE.

B. Random Walk With Restart

RWR model has been proven to be effective in many network-based recommendation systems [47], [48]. RWR provides an effective way to estimate the similarity between two nodes in a graph. Given a network and a target node, the basic random walk (RW) model selects a neighbor of the target node randomly, and moves to this neighbor; then the RWR model selects a neighbor of this node at random, and moves to it. The whole random sequence of node selection is a random walk on graph [49]. However, the basic RW model may leave the target node quickly, resulting the loss of the context. The RWR model is proposed to avoid such situation. In RWR model, in each transition, there is a random probability of jumping back to the original node.

The RWR model has received extensive interest from both practical and theoretical points of view. There are several reasons behind the success of RWR. The main reason is that the RWR model takes advantages of both the node and network structure information simultaneously without losing the information. Meanwhile, the random walk process can be guided to the more relevant nodes by considering the link weights.

Konstas *et al.* [50] design a collaboration recommendation system based on the generic framework of RWR in order to provide a more accurate and efficient way to represent social networks. Through extensive experiments, they find that the RWR model outperforms the standard collaborative filtering method since it utilizes the additional information embedded in the social knowledge. Zhou *et al.* [51] design a collaborator recommendation system in heterogeneous network using RWR model. In contrast to the previous random walk on homogeneous network, they construct a heterogeneous network with multiple types of nodes and links with a general network structure by excluding the citing paper nodes. Then, they bias the random walker with weighted links. Due to the effectiveness of RWR model, we adopt it as the basic framework of SCORE model.

C. Diversified Recommendation

Previous recommendation systems mainly recommend items similar to the items that target users are interested in. From a long-term perspective, users may be bored with the recommended items because they may already have discovered such items. In order to deal such issues, recommendation system should provide diverse suggestions, where the diversity should be considered. The recommendation system with poor diversity or quality might frustrate users [9], [25], [52]. Most recommendation systems are mainly evaluated based on accuracy, which may narrow down the users' horizons. Moreover, the novelty and diversity of the suggested items are overlooked. This may lead to a low user satisfaction.

Recently, scholars have paid more attentions to the diversity issue in recommendation systems [53]. The recommendation diversity can be analyzed either at an aggregate level or an individual level. In aggregate diversity, the recommendation system can be evaluated by the slope of the long-linear relationships between item popularity and recommendations [54]. Adomavicius and Kwon [55] propose a series of ranking techniques that have higher aggregate diversity across all users, while maintaining comparable levels of recommendation accuracy. The individual diversity is personalized to each user, which tries to recommend personalized items to target users. The topic coverage has been used as a metric to evaluate the individual diversity [56].

In this paper, we aim to recommend sustainable collaborators to target scholars. We consider not only the similarity between scholar pairs but also the weak tie relationships between them for the purpose of SCORE. Therefore, our recommended collaborators are more diverse than traditional recommendation system from the network perspective. The diversity in our recommendation system is the sustainability. To reach sustainable recommendation, we have to recommend more diverse collaborators. Specifically, we utilize the conference attending information to enhance the individual diversity of collaborator recommendation.

VI. Conclusion

This paper proposed to recommend sustainable collaborations for scholars by taking advantages of the conference closure. The conference closure was utilized to bias the random walk in scientific collaboration networks to recommend diverse potential collaborators. Through extensive experiments on data sets extracted from DBLP with attendees of ten academic conferences, we demonstrated the effectiveness of proposed model SCORE. The SCORE could recommend sustainable collaborators whilst limiting the decrease of accuracy. The results of this paper indicated that when doing collaborator recommendations, we should not only consider the network similarity, i.e., common neighbors but also additional weak ties, i.e., conference coattending information. The findings have significant implications for understanding the social functions of academic conferences.

Despite its exploratory nature, this paper offers some insights into scientific collaborator recommendation from the perspective of sustainability. In the future, we will explore more about the relationships between academic conferences and scientific collaborations. Meanwhile, with the help of SCORE model, we will try to provide a real online recommendation system for collaborator recommendations to overcome the information overload problems in the age of scholarly big data.

Acknowledgment

The authors would like to thank Y. Yuan for the valuable discussions on the experimental design. They would also like to thank all the reviewers for their efforts in improving the quality of this paper.

References

[1] L. Waltman, R. J. W. Tijssen, and N. J. V. Eck, "Globalisation of science in kilometres," *J. Informetrics*, vol. 5, no. 4, pp. 574–582, Oct. 2011.

[2] R. J. W. Tijssen, L. Waltman, and N. J. V. Eck, "Collaborations span 1,553 kilometres," *Nature*, vol. 473, no. 7346, p. 154, May 2011.

[3] J. Adams, "The fourth age of research," *Nature*, vol. 497, no. 7451, pp. 557–560, May 2013.

[4] S. Khan, X. Liu, K. A. Shakil, and M. Alam, "A survey on scholarly data: From big data perspective," *Inf. Process. Manage.*, vol. 53, no. 4, pp. 923–944, Jul. 2017.

[5] C. Zhang, Y. Bu, Y. Ding, and J. Xu, "Understanding scientific collaboration: Homophily, transitivity, and preferential attachment," *J. Assoc. Inf. Sci. Technol.*, vol. 69, no. 1, pp. 72–86, Jan. 2018.

[6] W. Wang, S. Yu, T. M. Bekele, X. Kong, and F. Xia, "Scientific collaboration patterns vary with scholars' academic ages," *Scientometrics*, vol. 112, no. 1, pp. 329–343, Jul. 2017.

[7] D. Liben-Nowell and J. Kleinberg, "The link-prediction problem for social networks," *J. Assoc. Inf. Sci. Technol.*, vol. 58, no. 7, pp. 1019–1031, May 2007.

[8] F. Xia, Z. Chen, W. Wang, J. Li, and L. T. Yang, "MVCWalker: Random walk-based most valuable collaborators recommendation exploiting academic factors," *IEEE Trans. Emerg. Topics Comput.*, vol. 2, no. 3, pp. 364–375, Sep. 2014.

[9] P. Cheng, S. Wang, J. Ma, J. Sun, and H. Xiong, "Learning to recommend accurate and diverse items," in *Proc. 26th Int. Conf. World Wide Web*, 2017, pp. 183–192.

[10] A. M. Petersen, "Quantifying the impact of weak, strong, and super ties in scientific careers," *Proc. Nat. Acad. Sci. USA*, vol. 112, no. 34, pp. E4671–E4680, Aug. 2015.

[11] A. Rapoport, "Spread of information through a population with sociostructural bias: I. Assumption of transitivity," *Bull. Math. Biophys.*, vol. 15, no. 4, pp. 523–533, Dec. 1953.

[12] G. Kossinets and D. J. Watts, "Empirical analysis of an evolving social network," *Science*, vol. 311, no. 5757, pp. 88–90, Jan. 2006.

[13] M. S. Granovetter, "The strength of weak ties," *Amer. J. Sociol.*, vol. 78, no. 6, pp. 1360–1380, May 1973.

[14] X. Su, W. Wang, S. Yu, C. Zhang, T. M. Bekele, and F. Xia, "Can academic conferences promote research collaboration?" in *Proc. 16th ACM/IEEE-CS Joint Conf. Digit. Libraries*, Jun. 2016, pp. 231–232.

[15] W. Wang, X. Bai, F. Xia, T. M. Bekele, X. Su, and A. Tolba, "From triadic closure to conference closure: The role of academic conferences in promoting scientific collaborations," *Scientometrics*, vol. 113, no. 1, pp. 177–193, Oct. 2017.

[16] R. M. Roberts, *Serendipity: Accidental Discoveries in Science*. Hoboken, NJ, USA: Wiley, 1989, p. 288.

[17] H. Tong, C. Faloutsos, and J.-Y. Pan, "Random walk with restart: Fast solutions and applications," *Knowl. Inf. Syst.*, vol. 14, no. 3, pp. 327–346, Mar. 2008.

[18] W. Wang, Z. Cui, T. Gao, S. Yu, X. Kong, and F. Xia, "Is scientific collaboration sustainability predictable?" in *Proc. 26th Int. Conf. World Wide Web Companion*, 2017, pp. 853–854.

[19] J. E. Hirsch, "An index to quantify an individual's scientific research output," *Proc. Nat. Acad. Sci. USA*, vol. 102, no. 46, p. 16569–16572, Nov. 2005.

[20] G. Bianconi, R. K. Darst, J. Iacovacci, and S. Fortunato, "Triadic closure as a basic generating mechanism of communities in complex networks," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 90, no. 4, Oct. 2014, Art. no. 042806.

[21] A. Töscher, M. Jahrer, and R. Legenstein, "Improved neighborhood-based algorithms for large-scale recommender systems," in *Proc. 2nd KDD Workshop Large-Scale Recommender Syst. Netflix Prize Competition*, 2008, p. 4.

[22] D. Easley and J. Kleinberg, *Networks, Crowds, and Markets: Reasoning about a Highly Connected World*. Cambridge, U.K.: Cambridge Univ. Press, 2010.

[23] M. McPherson, L. Smith-Lovin, and J. M. Cook, "Birds of a feather: Homophily in social networks," *Annu. Rev. Sociol.*, vol. 27, no. 1, pp. 415–444, Aug. 2001.

[24] R. S. Burt, "Structural holes and good ideas," *Amer. J. Sociol.*, vol. 110, no. 2, pp. 349–399, Sep. 2004.

[25] D. Kotkov, S. Wang, and J. Veijalainen, "A survey of serendipity in recommender systems," *Knowl.-Based Syst.*, vol. 111, pp. 180–192, Nov. 2016.

[26] L. Lü and T. Zhou, "Link prediction in complex networks: A survey," *Phys. A, Statist. Mech. Appl.*, vol. 390, no. 6, pp. 1150–1170, Mar. 2011.

[27] J. Li, F. Xia, W. Wang, Z. Chen, N. Y. Asabere, and H. Jiang, "ACRec: A co-authorship based random walk model for academic collaboration recommendation," in *Proc. 23rd Int. Conf. World Wide Web*, 2014, pp. 1209–1214.

[28] F. Xia, W. Wang, T. M. Bekele, and H. Liu, "Big scholarly data: A survey," *IEEE Trans. Big Data*, vol. 3, no. 1, pp. 18–35, Mar. 2017.

[29] A. Gunawardana and G. Shani, "A survey of accuracy evaluation metrics of recommendation tasks," *J. Mach. Learn. Res.*, vol. 10, pp. 2935–2962, Dec. 2009.

[30] X. Zhang, X. Zhang, H. Liu, and X. Liu, "Multi-task multi-view clustering," *IEEE Trans. Knowl. Data Eng.*, vol. 28, no. 12, pp. 3324–3338, Dec. 2016.

[31] J. Tang, J. Zhang, L. Yao, J. Li, L. Zhang, and Z. Su, "Arnetminer: Extraction and mining of academic social networks," in *Proc. 14th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2008, pp. 990–998.

[32] J. Tang, "Aminer: Toward understanding big scholar data," in *Proc. Ninth ACM Int. Conf. Web Search Data Mining*. Feb. 2016, pp. 467–468.

[33] X. Kong, H. Jiang, Z. Yang, Z. Xu, F. Xia, and A. Tolba, "Exploiting publication contents and collaboration networks for collaborator recommendation," *PLoS ONE*, vol. 11, no. 2, Feb. 2016, Art. no. e0148492.

[34] D. Surian, N. Liu, D. Lo, H. Tong, E.-P. Lim, and C. Faloutsos, "Recommending people in developers' collaboration network," in *Proc. 18th Work. Conf. Rev. Eng.*, Oct. 2011, pp. 379–388.

[35] S. Pramanik, M. Gundapuneni, S. Pathak, and B. Mitra, "Can i foresee the success of my meetup group?" in *Proc. IEEE/ACM Int. Conf. Adv. Social Netw. Anal. Mining (ASONAM)*. 2016, pp. 366–373.

[36] J. Zhang, "Uncovering mechanisms of co-authorship evolution by multirelations-based link prediction," *Inf. Process. Manage.*, vol. 53, no. 1, pp. 42–51, Jan. 2017.

[37] W. Zheng, Z. Yao, Y. Xie, J. Fan, and B. Hu, "Identification of Alzheimer's disease and mild cognitive impairment using networks constructed based on multiple morphological brain features," *Biolog. Psychiatry, Cognit. Neurosci. Neuroimag.*, vol. 3, no. 10, pp. 887–897, Oct. 2018.

[38] X. Zhang, C. Wang, Y. Su, L. Pan, and H.-F. Zhang, "A fast overlapping community detection algorithm based on weak cliques for large-scale networks," *IEEE Trans. Comput. Social Syst.*, vol. 4, no. 4, pp. 218–230, Dec. 2017.

[39] Z. Liu, X. Xie, and L. Chen, "Context-aware academic collaborator recommendation," in *Proc. 24th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Jul.2018, pp. 1870–1879.

[40] X. Kong, M. Mao, W. Wang, J. Liu, and B. Xu, "VOPRec: Vector representation learning of papers with text information and structural identity for recommendation," *IEEE Trans. Emerg. Topics Comput.*, to be published. doi: 10.1109/TETC.2018.2830698.

[41] H. H. Chen, L. Gou, X. Zhang, and C. L. Giles, "Collabseer: A search engine for collaboration discovery," in *Proc. ACM/IEEE Joint Conf. Digit. Libraries*, Jul. 2011, pp. 231–240.

[42] Y. Zhang, C. Zhang, and X. Liu, "Dynamic scholarly collaborator recommendation via competitive multi-agent reinforcement learning," in *Proc. 11th ACM Conf. Recommender Syst.*, Aug. 2017, pp. 331–335.

[43] X. Bai, M. Wang, I. Lee, Z. Yang, X. Kong, and F. Xia, "Scientific paper recommendation: A survey," *IEEE Access*, vol. 7, pp. 9324–9339, 2019.

[44] C.-H. Tsai and Y.-R. Lin, "Tracing and predicting collaboration for junior scholars," in *Proc. 25th Int. Conf. Companion World Wide Web*, 2016, pp. 375–380.

[45] Y. Sun, J. Han, C. C. Aggarwal, and N. V. Chawla, "When will it happen?: Relationship prediction in heterogeneous information networks," in *Proc. 5th ACM Int. conf. Web Search Data Mining*, 2012, pp. 663–672.

[46] J. Tang, S. Wu, J. Sun, and H. Su, "Cross-domain collaboration recommendation," in *Proc. 18th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2012, pp. 1285–1293.

[47] H. Bagci and P. Karagoz, "Context-aware location recommendation by using a random walk-based approach," *Knowl. Inf. Syst.*, vol. 47, no. 2, pp. 241–260, May 2016.

[48] J. Liu *et al.*, "Artificial intelligence in the 21st century," *IEEE Access*, vol. 6, pp. 34403–34421, 2018.

[49] L. Lovász, "Random walks on graphs," *Combinatorics, Paul Erdos Is Eighty*, vol. 2. Berlin, Germany: Springer, 1993, pp. 1–46.

[50] I. Konstas, V. Stathopoulos, and J. M. Jose, "On social networks and collaborative recommendation," in *Proc. 32nd Int. ACM SIGIR Conf. Res. Develop Inf. Retr.* 2009, pp. 195–202.

[51] X. Zhou, L. Ding, Z. Li, and R. Wan, "Collaborator recommendation in heterogeneous bibliographic networks using random walks," *Inf. Retr. J.*, vol. 20, no. 4, pp. 317–337, Aug. 2017.

[52] L. Cai *et al.*, "Scholarly impact assessment: A survey of citation weighting solutions," *Scientometrics*, vol. 118, no. 2, pp. 453–478, Dec. 2018. doi: 10.1007/s11192-018-2973-6.

[53] A. Ashkan, B. Kveton, S. Berkovsky, and Z. Wen, "Optimal greedy diversity for recommendation," in *Proc. IJCAI*, 2015, pp. 1742–1748.

[54] S. M. McNee, J. Riedl, and J. A. Konstan, "Being accurate is not enough: How accuracy metrics have hurt recommender systems," in *Proc. Extended Abstr. Hum. Factors Comput. Syst. (CHI)*, 2006, pp. 1097–1101.

[55] G. Adomavicius and Y. Kwon, "Improving aggregate recommendation diversity using ranking-based techniques," *IEEE Trans. Knowl. Data Eng.*, vol. 24, no. 5, pp. 896–911, May 2012.

[56] S. Vargas, L. Baltrunas, A. Karatzoglou, and P. Castells, "Coverage, redundancy and size-awareness in genre diversity for recommender systems," in *Proc. 8th ACM Conf. Recommender Syst.*, 2014, pp. 209–216.

Wei Wang received the Ph.D. degree in software engineering from Dalian University of Technology, Dalian, China, in 2018.

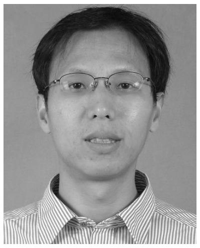
His current research interests include big scholarly data, social network analysis, and computational social science.

Jiaying Liu received the B.S. degree in software engineering from Dalian University of Technology, Dalian, China, in 2016, where she is currently pursuing the master's degree with the School of Software.

Her current research interests include big scholarly data, social network analysis, and science of success.

Zhuo Yang received the M.S. degree from Nanyang Technological University, Singapore, in 2007.

She is currently a Software Engineer with the School of Software, Dalian University of Technology, Dalian, China. Her current research interests include software auto-testing, software quality assurance, and project management.



Xiangjie Kong (M'13–SM'17) received the B.Sc. and Ph.D. degrees from Zhejiang University, Hangzhou, China.

He is currently an Associate Professor with the School of Software, Dalian University of Technology, Dalian, China. He has authored or co-authored over 100 scientific papers in international journals and conferences. His current research interests include computational social science, mobile computing, and big data.

Dr. Kong is a senior member of the CCF and a member of the ACM. He served as a (Guest) Editor of several international journals and the Workshop Chair or a PC Member of a number of conferences.

Feng Xia (M'07–SM'12) received the B.Sc. and Ph.D. degrees from Zhejiang University, Hangzhou, China.

He was a Research Fellow at the Queensland University of Technology, Brisbane, QLD, Australia. He is currently a Full Professor with the School of Software, Dalian University of Technology, Dalian, China. He has authored two books and over 200 scientific papers in international journals and conferences. His current research interests include data science, big data, knowledge management, network science, and mobile social networks.

Dr, Xia is a senior member of the ACM and a member of the AAAS.